A Neural Network Model of Object Segmentation and Feature Binding in Visual Cortex

N00014-90-J-1864

Paul Sajda and Leif H. Finkel
Departmen of Bioengineering and
Institute of Neurological Sciences
University of Pennsylvania
Philadelphia, PA. 19104-6392

1990

Abstract

We present neural network simulations of how the visual cortex may segment objects and bind attributes based on depth-from-occlusion. We briefly discuss one particular subprocess in our occlusion-based model most relevant to segmentation and binding: determination of the *direction of figure*. We propose that our model allows us to address a central issue in object recognition: how the visual system defines an object. In addition, we test our model on "illusory" stimuli, with the network's response indicating the existence of robust psychophysical properties in the system.

Introduction

In order to discriminate objects in the visual world, the nervous system must solve two fundamental problems: object segmentation and binding. Object segmentation deals with the problem of how separate objects are distinguished. Conversely, the binding problem [1], addresses how specific attributes—shape, color, motion, and depth—are linked to create an individual object. In a typical visual scene, multiple objects occlude one another, creating a perceptual dilemma—to which of the two overlapping surfaces does the common border belong? If the border is, in fact, an occlusion border, then it belongs to the occluding object. This identification results in a stratification of the two objects in depth and a de facto discrimination of the objects. For example, consider the case of a horse behind a tree. We perceive the tree as being closer than the horse, and in addition, the two "halves" of the horse created by the occlusion are linked into one object. Thus, though occlusion may isolate different regions of an object, our visual system is able to overcome this difficulty and provide a consistent and coherent representation of the scene's constituent objects.

We have developed a neural network model of how the visual cortex may discriminate objects using depth-from-occlusion—the process of determining depth relationships based solely on object interpostion. The following provides a brief overview of the model's organization and function. In addition, we discuss the neural circuitry of a particular subprocess in the system, which we call *direction of figure*—a critical operation for solving the problems of object segmentation and binding. Finally, we demonstrate, with simulations, how the system is able to segment objects and stratify them in depth. A more detailed description of the operation of the model is described elsewhere [4].

Overview of the model

A system must identify if an occlusion relationship exists if it is to accurately segment an image and determine relative depth. Since occlusion implies discontinuous depth, one can conclude that discontinuities in the image provide important occlusion information. Given that an object always occludes its background, all objects possess an *occluding contour* [9] in their two dimensional image. An occluding contour is a closed curve which "outlines" an object's silhouette. Though an occluding contour signals occlusion with the background, it alone gives little information about depth relationships between objects. In the two dimensional image there are a number of cues which imply object interposition. The strongest cue is the *T-junction*. At a T-junction the contours of occluding and occluded objects meet. T-junctions have long been recognized as important cues for scene segmentation [8]. Two other cues to occlusion are *concavities* and *surrounded contours*. Objects at different depths have overlapping two dimensional images, creating concavities in the occluding contours of the objects. The presence of concavities can therefore serve as an indicator for occlusion. Another occlusion cue occurs when a smaller object is in front of a larger object but no T–junctions are created. In this case the smaller object is completely surround by the larger object's occluding contour. This surround condition can be interpreted as a cue to occlusion, with the smaller object perceived to be closer to the viewer. However, since objects often contain concavities or surrounded contours (for example an annulus) as part of their structure, neither concavities nor surrounded contours are as strong a cue to occlusion as T–junctions.

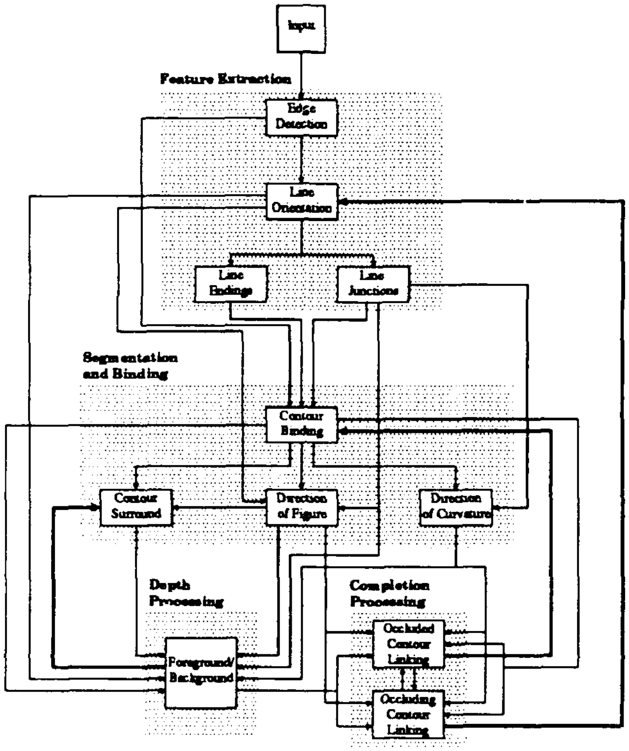

Figure 1: Organization and information flow in our occlusion–based model of object segmentation.

Our model identifies and uses these occlusion cues to segment the two dimensional image and determine the relative depth of objects. A schematic of information flow in the model is shown in figure 1. The system is organized into four main categories of processing; *feature extraction*, ***segmentation and binding***, *depth processing* and *completion processing*. These in turn are divided into subcategories which represent functions performed by particular networks in the system. All networks are organized topographically, consistent with observed organization of the visual cortex. In addition, reentrant feedback [3], found throughout the cortex, is used extensively to integrate information between different networks.

The first stage of the model discriminates low–level features. Edges, oriented lines, line endings and junctions are detected by networks of units selective for these image features. The next stage involves segmentation and binding, which includes grouping features into *proto–objects*. We define a proto–object as *a compact region, surrounded by a closed, piecewise continuous contour (occluding contour), and located at a certain depth*. Proto–objects are the precursors of objects, since feedback from completion processing (see figure 1) can group one or more proto–objects into a single object. The third stage of the model involves completing occluded and occluding contours. In our previous example, the two regions of the horse separated by the occluding tree are perceptually linked to form a single object.[1] This stage of the model includes mechanisms for linking unoccluded portions of proto–objects. In addition, occluding contours may be incomplete. For example, the luminance contrast between the tree and the horse may be small over a region of the image, and therefore no edge discontinuity is detectable. However, unambiguous

[1] Within our model, the two"halves" of the horse would be classified as proto–objects.

edges are linked within the model to form continuous closed occluding contours. The final stage of the model is concerned with depth processing. Here a cooperative/competitive mechanism uses occlusion cues, identified in the earlier stages of processing, as "forces" for "pushing" objects into different relative depths. Depth in our model is represented in a distributed fashion between units in *foreground* and *background* networks. This distributed representation of depth is consistent with how disparity is represented in visual cortex [12][14].

The model is simulated using the NEXUS Neural Simulator [15], a system designed for modelling multiple interconnected neural maps. Present simulations consist of 42 topographically organized networks, each containing 4096 units (64x64). The functions performed by the individual units range from simple linear thresholding to execution of more complex instructions and algorithms. We allow for this range in a unit's function by defining PGN (Programmable Generalized Neural) units for modelling the functionality of neural assemblies and larger neuronal circuits. No learning is involved in the network dynamics as the model is intended to correspond to preattentive perception.

Direction of figure

Networks in the feature extraction stage of the model are able to identify edges and determine orientation and curvature of contours. However, these units are not sufficient for segmenting the image into its constituent proto-objects. The problem also requires that the surface of the object be identified. The task of the direction of figure network is to determine which side of the contour is the "inside" (surface) of the object and which is the "outside" (background). The problem can be restated as determining which region "owns" the contour [4][11][13].

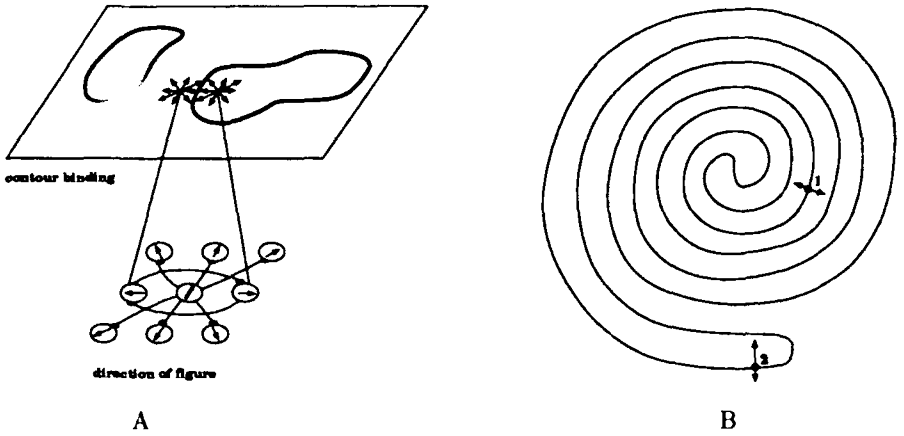

Figure 2: **A** Neural circuit for determining direction of figure (inside vs. outside), with a hypothetical input stimulus consists of two closed contours (bold curves). **B** An example of a stimulus where our proposed neural circuit cannot correctly determine direction of figure for the entire object.

A schematic of a neural mechanism for computing the direction of figure is shown in figure 2A. The function of this circuit is based on the following observation. Suppose a unit projects its dendrites (connections) in a stellate configuration to units in the *contour binding* network and that these dendrites are activated by units responding to a particular contour. We will not go into the details of the contour binding network's operation except to say that the activity in this network represents a "tag" for the individual occluding contours defining the proto-objects. Units which lie on the same contour are bound together with this common tag[2]. In the direction of figure network, if a given unit is inside a contour, more of its dendrites will be activated than if it is outside the contour. A winner-take-all mechanism between two such units will determine which is more strongly activated, and hence which is the inside of the object (in figure 2A, units representing possible directions are shown with arrows). As shown in figure 2A it is

[2] The biological substrate for such a binding mechanism may be cortical oscillations or phase-locking [5][2].

advantageous to limit this competition to the two units which are located at positions directly perpendicular to the local orientation of the contour. It is important to note that this mechanism is consistent with human perception in that it will fail to identify the correct direction of figure for selected cases (see figure 2B).

Simulation results

Figure 3A shows a scene that was presented to system. The low-level networks detect edges, line orientation, terminations and junctions present in the scene. Figure 3B displays the activity in the contour binding network, representing the tags assigned to the different scene elements. Each box represents elements having a common tag, different boxes represent different tags, and the ordering of the boxes is arbitrary. On the first cycle (see figure 3B *top*) discontinuous elements, such as the two regions of the horse, have separate tags. Feedback from the completion networks links these contours so that after the second cycle (see figure 3B *bottom*)the contours defining the horse have the same tag.

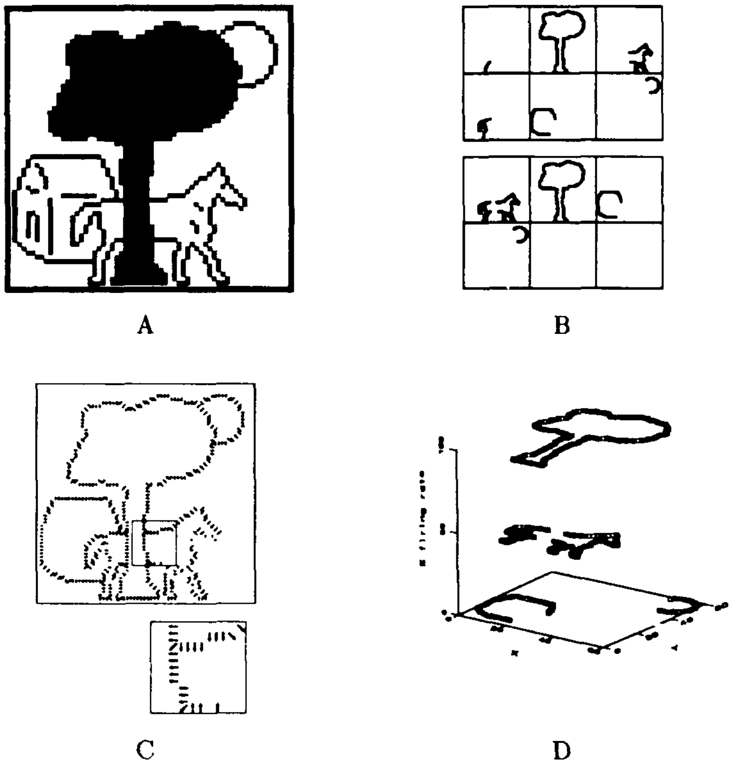

Figure 3: **A** A 64x64 stimulus presented to the system. **B** Spatial histogram of the contour binding tags (each box shows a unit with a common tag, different boxes represent different tags, and the ordering of the boxes is arbitrary). **C** Output of the direction of figure network after two cycles. Inset shows a magnified view of the output of the direction of figure network for a local section of the image. **D** Relative depth of objects in the scene as determined by the system.

The output of the direction of figure network for this particular stimulus is shown in figure 3C. The direction of the arrows indicates the direction of figure determined by the network. A small portion of the network is enlarged to better illustrate the system's performance. Note that the system correctly determines that the region representing the surface of the tree "owns" the vertical contour, while the surrounding contour is "owned" by the region of the horse.

T-junctions, such as those between the horse and the tree, force the various objects into different depth planes. The result of this process is shown in figure 3D, which plots the firing rate (as a percent of maximum) of units in the foreground network. The actual depth value determined for each object is somewhat arbitrary, and can vary depending upon minor changes in the scene–the system is designed to achieve the correct relative ordering, not absolute depth.

The second simulation illustrates that the system displays a response consistent with human responses to illusory stimuli. Figure 4A shows a stimulus known as the Kanizsa square [10]. Human subjects typically perceive a white square occluding four black discs and a wireframe square. This perception is somewhat surprising given that the stimulus can just as easily be interpreted as four black "pacman-like" shapes and four angular line segments. Some have suggested that the perception of these illusions may arise from artificially arranged occlusion cues [6].

Figure 4B shows the output of the direction of figure network after one and three cycles of activity. The large display shows that the surfaces of the objects (the discs, occluded and occluding squares) are correctly identified by the network after the third cycle. The two insets show an enlarged area of the network for both the first and third cycle. At first the system identifies the "L"-shaped mouth of the pacman as belonging to the disc, as illustrated by the direction of figure arrows. After the third cycle the "L"-shaped edge is identified as belonging to the occluding illusory square. This change in ownership of the edge results from the identification of occlusion—the edge has been identified as an occlusion border. Figure 4C displays the firing rate of units in the *foreground* depth map (as in figure 3D), thus showing that the system discriminates relative depth of the constituent objects.

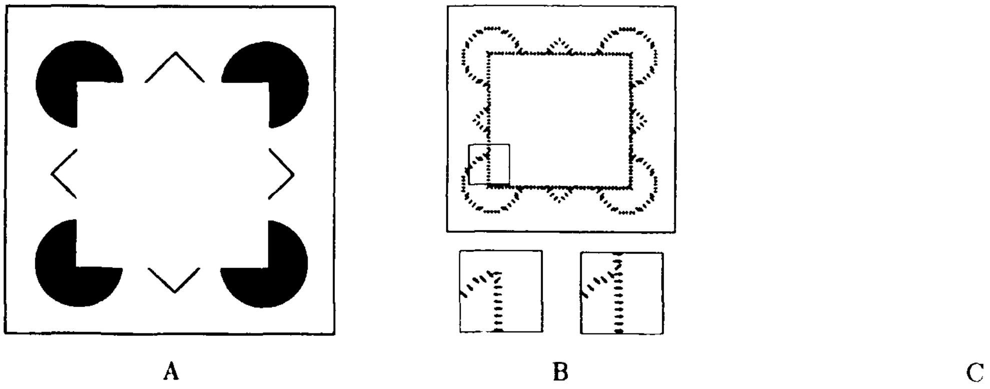

Figure 4: **A** Stimulus which is perceived by human subjects as an "illusory" square occluding four black discs and a wireframe square (rotated by 45°). A 64x64 discrete version of the stimulus was presented to the system. **B** Direction of figure determined by the system after three cycles. Insets show an enlarged view of a section of the output after the first and second cycle. **C** Activity in the foreground network (% maximum) demonstrating that the network correctly determines the relative depth for this illusory stimulus.

Conclusion

We have presented a neural network model of preattentive processing, capable of determining depth-from-occlusion, and have focused on those subprocesses within the model relevant to object segmentation and binding. We have also addressed issues for identifying the "ownership" of contours and how they relate to the figure/ground problem.

The model bears certain similarities to previous neural and psychological studies. Several models of illusory contour generation [3][7][16] have used related mechanisms to check for collinearity or generate illusory contours. Our model differs at a more fundamental level—we are concerned with objects not just contours. To define an object, surfaces must also be considered. In a simple line drawing, we perceive an interior surface despite the fact that no surface properties are indicated. Thus, the model must characterize the surface—and it does so by determining the direction of figure and relative depth. A more complete model will include additional surface properties such as color, brightness, texture, and surface orientation.

Acknowledgements

This research was supported by grants from The Office of Naval Research (N00014-90-J-1864), The Whitaker Foundation, and The McDonnell-Pew Program in Cognitive Neuroscience.

References

[1] H. B. Barlow. Critical limiting factors in the design of the eye and visual cortex. *Proc. Royal Society (London)*, B212:1-34, 1981.

[2] R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk, and H. Reitboeck. Coherent oscillations: a mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60:121–130, 1988.

[3] L. Finkel and G. Edelman. Integration of distributed cortical systems by reentry: a computer simulation of interactive functionally segregated visual areas. *Journal of Neuroscience*, 9:3188–3208, 1989.

[4] L.H. Finkel and P. Sajda. Object discrimination based on depth-from-occlusion. *Neural Computation*, submitted.

[5] C. M. Gray and W. Singer. Neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Science USA*, 86:1698-1702, 1989.

[6] R. L. Gregory. Cognitive contours. *Nature*, 238:51–52, 1972.

[7] S. Grossberg and E. Mingolla. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychology Review*, 92:173-211, 1985.

[8] A. Guzman. Decomposition of a visual scene into three-dimensional bodies. *Fall Joint Computer Conference*, 1968:291–304, 1968.

[9] D. Marr E. Hildreth. Analysis of occluding contour. *Proceedings of the Royal Society of London (Biology)*, 197:441–475, 1977.

[10] G. Kanizsa. *Organization in Vision*. Praeger, New York, 1979.

[11] K. Koffka. *Principles of Gestalt Psychology*. Harcourt, Brace, New York, 1935.

[12] S. Lehky and T. Sejnowski. Neural model of stereoacuity and depth interpolation based on distributed representation of stereo disparity. *Journal of Neuroscience*, 7:2281–2299, 1990.

[13] K. Nakayama and S. Shimojo. Toward a neural understanding of visual surface representation. *Cold Spring Harbor Symposia on Quantitative Biology*, LV:911–924, 1990.

[14] G. F. Poggio, F. Gonzalez, and F. Krause. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *Journal of Neuroscience*, 8:4531–4550, 1988.

[15] P. Sajda and L. Finkel. NEXUS: A simulation environment for large-scale neural systems. *SIMULATION*, submitted.

[16] S. Ullman. Filling-in the gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics*, 25:1–6, 1976.